



## SAVVY SEARCHING

Open access

# Open access to scholarly full-text documents

Peter Jacsó

*University of Hawaii, Hawaii, USA*

587

### Abstract

**Purpose** – The purpose of this article is to discuss open access to scholarly full-text documents.

**Design/methodology/approach** – Discusses open access to scholarly full-text documents.

**Findings** – The paper shows that while open access archives are good for the majority, for publishers, editors and authors, open access articles can substantially increase their impact, and the impact factor for the source journals.

**Originality/value** – The paper offers insights into scholarly full-text documents.

**Keywords** Digital storage, Universities, Publishers

**Paper type** General review

The last issue of this column discussed the open access alternatives to scholarly indexing/abstracting information (Jacsó, 2006a). These can fully or partially substitute for some fee-based indexing/abstracting publications in print, CD-ROM and/or online formats, and save thousands of dollars for libraries in identifying relevant published documents about a subject. For many libraries, however, the open access secondary services are of little help if the libraries do not subscribe to the print edition of the source journals covered by the secondary services, and do not have sufficient budget for interlibrary loans, let alone for document delivery services. This is a particularly acute problem in the developing countries. In addition, for many users resource discovery through indexing/abstracting publications is not gratifying, even if the primary documents are (almost) instantly available in or through the library in print format.

This column discusses the variety of open access alternatives to the full-text of primary scholarly documents in digital format. These in turn may offer not “only” the full-text of the documents but also supplements and additional contents (such as detailed versions of the summary tables, questionnaires, colour graphs, charts and other materials) which are not available in print format for technical, monetary or space reasons. Most importantly, the majority of the providers of full-text open access digital scholarly document collections provide the option of searching the full-text itself – although with varying degrees of efficiency.

### Providers and types of open access scholarly full-text documents

The largest collections of scholarly full-text documents are provided by digital depositories and repositories maintained by government agencies, associations, universities, professional volunteer groups, as well as new and traditional scholarly publishers (directly or indirectly through their digital facilitators).

Some government agencies excel in offering open access to tens of thousands of scientific and technical reports, such as the National Criminal Justice Reference Service



---

(NCJRS), the Transportation Research Information Services (TRIS), the InfoBridge and the ERIC databases from the Department of Justice, the Department of Transportation, the Department of Energy, and the Department of Education, respectively. It is unfortunate that the open access version of the ERIC database does not search the full-text of the 110,000 ERIC Digest documents. You may limit the search to items which have full-text, but you will not find documents which have the search term only in the full-text and not in the bibliographic data (title, abstract, descriptor, identifier, etc.). It should be a quick fix to bring out the best from the open access version of the full-text enhancements in the ERIC database.

The PubMed Central database of the National Institute of Health is far the largest provider of open access full-text journal articles in medicine and the life sciences. PubMed itself has been a mightily powerful open access indexing/abstracting database with links to open access articles. PubMed Central took this a giant step further by having close to 650,000 open access full-text scholarly articles in its depository; these articles are fully and powerfully searchable – not just linked to with a little savvy – using Google, Google Scholar, Yahoo! with the site:pubmed.nih.gov filter. They all crawl the PubMed Central site, although none of them completely. This is an enormous qualitative advantage over any version of MEDLINE. PubMed Central covers some journals (the *British Medical Journal*, for example) which do not deposit their full-text articles in PMC, and only links are provided to the open access full-text papers; but there are only 5,270 such documents. Surprisingly few articles gave substantial accolades to the remarkable achievement of NIH (Delamothe and Smith, 2001).

Universities have been instrumental not only in developing software tools for digital libraries and building collections, primarily by facilitating the deposit of the full-text versions of the publications of their faculty members, and also by organizing conferences and offering assistance to act as depositories and repositories of full-text scholarly publications in a (sub)discipline. The arxiv.org repository of scholarly papers in physics can be considered the mother of all open access collections. It has been developed and maintained by Cornell University, one of the bastions in the development of digital libraries. Papers in computer science, cognitive sciences, and quantitative biology have been added to this e-print archive of about 380,000 open access full-text documents. It is searchable through several other information services, including both subscription-based and toll-free services, such as ISI Web Citation Index and Scirus, respectively.

One of the most interesting use of the arxiv.org collection (and a few other archives) is through CiteBase (Brody, 2003) – an experimental and incomplete but nevertheless impressive open access citation searching service developed by Tim Brody at Southampton University, a leader in developments and research projects related to the creation and use of open access collections and citation analysis in the UK.

One of the most successful projects to create a large, discipline-specific open access collection is CiteSeer (also known as ResearchIndex), which includes 750,000 computer and information science publications enhanced by a sophisticated autonomous citation searching component (Giles *et al.*, 1998). It was developed by NEC Laboratories and is maintained by Penn State University, which launched a similar open access full-text database, SMEALSearch, with a focus on open access business-related scholarly papers.

CiteSeer gave the idea to Google for the development of Google Scholar, and Google received unprecedented support from most publishers, who offered unfettered access to Google robots to harvest metadata and the full-text of millions of articles. While Google Scholar is a good tool for finding scholarly articles on practically any subject,

---

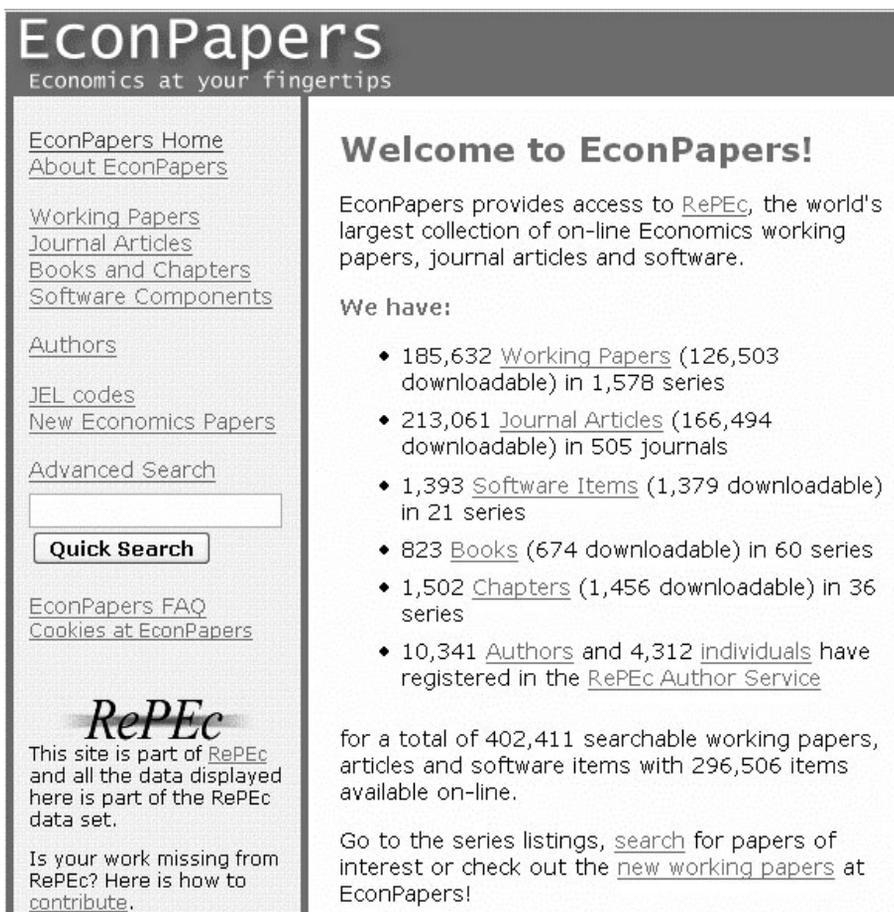
its poor structuring and identification of the data elements, and basic problems with the search engine, make it a disappointing tool for its much touted citation analysis function. Hiring a professional who is competent in developing algorithms for correctly matching the cited and citing references could provide a much better tool instead of dispensing highly inflated hit counts and absurd citation counts based on phantom citations that many researchers assume to be true (Jacsó, 2006c).

Scholarly associations and societies have been pioneering the open access movement, especially the American Association For Cancer Research, the American Cancer Society, the American Diabetes Association, the American Heart Association, the American Physiological Society, the American Meteorological Society, the American Society for Biochemistry and Molecular Biology, the American Society for Microbiology, the American Society for Pharmacology and Experimental Therapeutics, The Endocrine Society, the Society for Endocrinology, the Society for General Microbiology, and the Royal College of Psychiatrists which offer open access to hundreds of thousands full-text documents.

Professional volunteer groups have been very active in creating new, Web-born journals, converting existing print journals into digital serials, and/or creating open access depositories of scholarly publications. By far the largest and most impressive project by professional volunteers from 54 countries and 34 USA states is the RePEc (Research Papers in Economics) Project. This is a combination of personal and corporate directory of individuals and organisations doing economics research, a software directory and full-text archive of about 300,000 scholarly journal articles, conference papers, working papers (a key document type in economics), books and book chapters. The system has developed at such a fast rate that the home page reports almost 100,000 fewer documents than are really searchable and downloadable online. The availability of the metadata and the full-text documents in a very well-structured format, inspired a number of implementations by volunteers who focused on different services (source document searching, citation searching, etc.) and developed a variety of smart software features (see Figure 1).

Most of the other sources in this category of collectively created sources by information professionals have relatively small collections, but they are focused on specific topics within a discipline, such as the articles published *Ariadne*, *D-Lib Magazine*, *First Monday* on issues of digital libraries and related issues. In a few cases the competence, interest and stamina of a single individual produced remarkable results. The Web-born journal, *Information Research*, is the labor of love of the editor, Professor Tom Wilson of Sheffield University. It gives you a good indication of the quality of this journal that it is the only Web-born and open access journal in the Information and Library Science category of the Web of Science database, and it has a higher impact factor than 35 other journals in the category, including many traditional library and information science journals (some of which do not have any substantial digital presence on the Web, not even for subscribers).

Of the new publishers, BioMed Central stands out by offering open access to full-text documents in more than 160 journals. (Some of these are non-BMC journals which switched to using the BMC platform, such as the *Acta Veterinaria Scandinavica*, and became fully open access.) In most of the journals published or hosted by BMC all the research articles are open access, but there are significant exceptions for other types of publications, such as reviews and commissioned articles. For example, in *Arthritis Research & Therapy* there are 445 open access research articles, and 1,720 other articles. Many items in the latter category are also open access, but not all of



# EconPapers

Economics at your fingertips

[EconPapers Home](#)  
[About EconPapers](#)

[Working Papers](#)  
[Journal Articles](#)  
[Books and Chapters](#)  
[Software Components](#)

[Authors](#)

[JEL codes](#)  
[New Economics Papers](#)

[Advanced Search](#)

**Quick Search**

[EconPapers FAQ](#)  
[Cookies at EconPapers](#)

## RePEc

This site is part of [RePEc](#) and all the data displayed here is part of the RePEc data set.

Is your work missing from RePEc? Here is how to [contribute](#).

## Welcome to EconPapers!

EconPapers provides access to [RePEc](#), the world's largest collection of on-line Economics working papers, journal articles and software.

**We have:**

- 185,632 [Working Papers](#) (126,503 downloadable) in 1,578 series
- 213,061 [Journal Articles](#) (166,494 downloadable) in 505 journals
- 1,393 [Software Items](#) (1,379 downloadable) in 21 series
- 823 [Books](#) (674 downloadable) in 60 series
- 1,502 [Chapters](#) (1,456 downloadable) in 36 series
- 10,341 [Authors](#) and 4,312 [individuals](#) have registered in the [RePEc Author Service](#)

for a total of 402,411 searchable working papers, articles and software items with 296,506 items available on-line.

Go to the series listings, [search](#) for papers of interest or check out the [new working papers](#) at EconPapers!

**Figure 1.**  
Document types and  
volumes in the  
EconPapers  
implementation of RePEc

them. Limiting the search to 2006, the ratio in this journal was 90 research articles, and 27 other documents, of which 16 required subscription. In the entire collection there are about 17,000 research articles and 7,800 other articles. There is no way to exclude from the search the subscription-based items. Even with this restriction, it is yet another significant open access database.

Traditional scholarly publishers (for-profit and non-profit alike) and their digital facilitators appeared at the end of my list above, even though for many users they seem to have the largest number of “free” full-text documents. However, the major part of their full-text collections are free only for patrons of libraries which subscribe to the print journals and conference proceedings; that is, these are not really open access because they are not available for anyone from anywhere. For example, the largest scholarly publisher, Elsevier, has about 6.5 million full-text journal articles in its ScienceDirect digital collection, but they are free only through libraries which subscribe to the print edition of the specific journals of Elsevier and its imprints (such as Pergamon Press, Academic Press, etc.). True, there are a few temporarily open

access issues of some of the journals, but these are not predictable. To their credit, many of the traditional journals proved their generosity by offering open access to many of their journals to developing countries through the HINARI and AGORA projects (Kirsop and Chan, 2005).

That said, some of the traditional publishers do offer genuinely and universally open access to publications, and not only for less prominent journals, but also for ones with the highest impact in the discipline. Oxford University Press stand out among the traditional publishers both in the quantity and quality of open access journals. For example, there are more than 6,000 open access articles from *Human Reproduction*, *Human Reproduction Update* and *Molecular Human Reproduction* which are ranked first, second and fourth by impact factor among the 24 journals in the Reproductive Biology section of the most current ISI Journal Citation Reports.

HighWire Press provides the largest number of open access articles, about 1,370,000 articles as of the end of July 2006. This is close to 40 per cent of the 3.6 million full-text article collection of HighWire Press. Although Ingenta, MetaPress and Allen Press also host some open access journals for traditional publishers, HighWire Press is the largest and most outstanding digital facilitator (see Figure 2).

HighWire hosts nearly 200 of the journals of Oxford University Press alone (25 per cent of them open access), among the nearly 1000 journals of dozens of publishers. It hosts one third of the 200 highest impact factor journals in the sciences and social sciences. The third and fifth ranked journals of the Reproductive Biology Section, *Biology of Reproduction* and *Reproduction*, are also offered in open access format by HighWire Press on behalf of their respective publishers. It is the host also of *Proceedings of the National Academy of Sciences (PNAS)* with more than 28,500 open

ISI Web of Knowledge<sup>SM</sup> Journal Citation Reports

Journal Citation Reports<sup>®</sup>

WELCOME HELP MARKED LIST 2005 JCR Science Edition

Journal Title Changes

Journal Summary List

Journals from: subject categories REPRODUCTIVE BIOLOGY VIEW CATEGORY SUMMARY LIST

Sorted by: Impact Factor SORT AGAIN

Journals 1 - 20 (of 24) Page 1 of 2

MARK ALL UPDATE MARKED LIST Ranking is based on your journal and sort selections.

Mark	Rank	Abbreviated Journal Title (linked to journal information)	ISSN	Total Cites	Impact Factor	Immediacy Index	Articles	Cited Half-life
<input type="checkbox"/>	1	<a href="#">HUM REPROD UPDATE</a>	1355-4786	2239	5.449	0.767	43	4.8
<input type="checkbox"/>	2	<a href="#">HUM REPROD</a>	0268-1161	19646	3.669	0.693	527	6.0
<input type="checkbox"/>	3	<a href="#">BIOL REPROD</a>	0006-3363	19517	3.583	0.755	327	6.2
<input type="checkbox"/>	4	<a href="#">MOL HUM REPROD</a>	1360-9947	3772	3.191	0.427	117	5.2
<input type="checkbox"/>	5	<a href="#">REPRODUCTION</a>	1470-1626	2622	3.136	0.379	169	3.3

**Figure 2.** The five highest impact factor journals in Reproductive Biology are hosted by HighWire Press digital facilitator on behalf of the publishers

access full-text documents from this journal along with a plethora of free powerful and appealing add-on services. *PNAS* is the third highest impact factor serial publication in the Multidisciplinary category of the ISI Journal Citation reports.

It illustrates the volume and quality of the open access journals hosted by HighWire Press that the search on the keyword, "reproduction", found 117,000 articles. Of the first 150 hits checked 46 were open access full-text articles from top ranking journals (see Figure 3).

There are many open access scholarly journal article and conference paper collections being built around the world. They represent not merely a broad topical and geographical spectrum, but also prove the viability of implementing open access archives even in low income countries. It is no surprise that the USA, UK, Canada and Japan are in the forefront of the open access centralised and decentralised archiving of scholarly publications. There are significant open access archives also in continental Europe (Dobratz and Matthaei, 2003), and some of the low income countries (especially India and Brazil) have already proven that the open access movement will not be restricted to the G8 countries, but will spread to less developed regions. Some examples are discussed as feasible models in the keynote address of the Asia and Oceania section of the 2006 IFLA Conference (Jacso, 2006b)

### The extent of open access

The Web-born journals typically offer open access to all the papers published throughout their lifetime. The publishers who digitised their back issues have many

http://www.amjbot.org/cgi/content/abstract/87/11/1599#otherarticles

Google Search

HOME HELP FEEDBACK SUBSCRIPTIONS ARCHIVE SEARCH TABLE OF CONTENTS

(*American Journal of Botany*, 2000, 87:1599-1608.)  
© 2000 Botanical Society of America, Inc.

## The cost of realized sexual reproduction: assessing patterns of reproductive allocation and sporophyte abortion in a desert moss<sup>1</sup>

Lloyd R. Stark<sup>5,2</sup>, Brent D. Mishler<sup>3</sup> and D. Nicholas McLetchie<sup>4</sup>

<sup>2</sup> Department of Biological Sciences, University of Nevada, 4505 Maryland Parkway, Box 454004, Las Vegas, Nevada 89154-4004 USA; <sup>3</sup> University Herbarium, Jepson Herbarium, and Department of Integrative Biology, 1001 Valley Life Sciences Bld., #2465, University of California, Berkeley, California, 94720-2465 USA; and <sup>4</sup> T.H. Morgan School of Biological Sciences, 101 Morgan Bld., University of Kentucky, Lexington, Kentucky 40506-0225 USA

The desert moss *Syntrichia caninervis* exhibits one of the most skewed sex ratios in the plant kingdom, with female individuals far outnumbering male individuals (exceeding 14♀:1♂). The "cost of sex hypothesis" derives from allocational theory and predicts that the sex which is most expensive should be the rarer sex. This hypothesis, which, as considered here represents the realized cost of sexual reproduction, is contingent upon two assumptions that are explored: (1) that male sex expression is more expensive than female sex expression, and (2) that sexual reproduction is resource limited. Using inflorescence biomass and discounting sperm, male sex expression was found to be in the neighborhood of one order of magnitude more expensive than female sex expression, and this difference is reflected in higher numbers of gametangia per male inflorescence, presence of paraphyses in male inflorescences, and a much longer developmental time for male

**This Article**

- Full Text
- Full Text (PDF)
- Submit a response
- Alert me when this article is cited
- Alert me when eLetters are posted
- Alert me if a correction is posted

**Services**

- Email this article to a friend
- Similar articles in this journal
- Similar articles in ISI Web of Science
- Similar articles in PubMed
- Alert me to new issues of the journal
- Download to citation manager
- Cited by other online articles
- Search for citing articles in: ISI Web of Science (26)

**Google Scholar**

- Articles by Stark, L. R.
- Articles by McLetchie, D. N.
- Articles citing this Article

**PubMed**

- PubMed Citation
- Articles by Stark, L. R.
- Articles by McLetchie, D. N.

**Agricola**

- Articles by Stark, L. R.
- Articles by McLetchie, D. N.

**Figure 3.**  
Several add-on services  
available through  
HighWire Press

different strategies. The most convenient (but not necessarily essential) is when all the articles from Volume 1 Issue 1 until the current (or even forthcoming) issue(s) are open access. This is the case, for example, with almost all of the journals in the BMC portfolio except for non-research articles in three journals in addition to *Arthritis Research & Therapy* discussed above. Most of these started in 2001 or later. The same is true for the small but ever growing PLoS (Public Library of Science) journal family.

There are some other journals, especially in PubMed Central and Allen Press, which offer open access to the whole run of some journals, such as the *Journal of the Medical Library Association* (and its former title: *Bulletin of the Medical Library Association*) going back to 1911, *Medical History* from 1957, or the *Journal of Clinical Investigation* from 1924 (which is also available as open access journal on the HighWire Press platform, but only from 1996 onward). It is a nice gesture that HWP flashes a message that PMC offers the whole run of this important journal.

Most of the time there is another restriction for the start of open access coverage. This applies even to subscribers who had subscription to the print version for decades. For example, about 90 per cent of the Oxford University Press journals are available in open access format from 1996. Even subscribers to the print journals for the earlier years would need to have a separate digital subscription to the historical archive of those journals. At the exhibit of the Medical Library Exhibition I saw the concept of the historical archive being introduced by many publishers – all for the pre-1996 period. It remains to be seen if this will be a moving wall, i.e. next year it will change to pre-1997, etc.

Then there is another limit to the extent of open access by most of the publishers (again, there are the exceptions for the BMC and PLOS journals and a few other). This is the imposed delay in making the volumes available for several months or years after publication. This moratorium can range from three to 36 months. The most common moratorium seems to be one to two years. This is understandable, as not even the most generous publishers want to give away the family jewels, the most current issues. For a reality check, very often end-users learn about interesting articles only when they start appearing in abstracting/indexing databases or start being cited, which rarely happen sooner than several months after publication.

### **Good for the goose, good for the gander**

One can complain about these restrictions, but it is much smarter to learn about the existing and impressive open access options which can save a lot of money, sparing considerable document delivery and interlibrary loan expenses. Unlike HighWire Press, Document delivery services fail to alert the searcher to the fact that an article is available in open access format. Infotrieve, for example, happily charges \$12 handling charges plus a copyright fee for articles readily available in open access full-text collections. A current case study (Papin-Ramcharan and Dawe, 2006) discusses how the University of the West Indies managed to maintain good quality information services for its researchers in spite of the shrinking budget, relying heavily on open access scholarly databases.

Open access archives are good for the goose and good for the gander. The advantages for the end-users are obvious, but the advantages for the publishers, the editors and the authors have been less apparent until a short but illuminating article (Lawrence, 2001) demonstrated that open access of articles can substantially increase their impact, and implicitly, the impact factor of the source journals. Since that time many studies have confirmed (and a few questioned) the validity of these findings. Particularly interesting is

the report (Harnad and Brody, 2004) which compared the impact of open access and non open access articles from the same journals. Members of the OpCit Project at Southampton University have created an up-to-date, annotated Webliography (OpCit, 2006) about the effect of open access on the impact of articles, and conference papers.

The open access movement has achieved significant results in a few years. These achievements will foster the broadening of its scope, and the creation of new full-text document archives with enhanced software features.

---

### References

- Brody, T. (2003), "CiteBase search: autonomous citation database for e-print archives", available at: <http://eprints.ecs.soton.ac.uk/10677/>
- Delamothe, C. and Smith, R. (2001), "PubMed Central: creating an Aladdin's cave of ideas – we have seen the future, and it works", *BMJ*, Vol. 322 No. 7277, pp. 1-2, available at: [www.pubmedcentral.gov/articlerender.fcgi?tool=pmcentrez&artid=1119296](http://www.pubmedcentral.gov/articlerender.fcgi?tool=pmcentrez&artid=1119296)
- Dobratz, S. and Matthaei, B. (2003), "Open archives activities and experiences in Europe", *D-Lib Magazine*, Vol 9 No. 1, available at: [www.dlib.org/dlib/january03/dobratz/01dobratz.html](http://www.dlib.org/dlib/january03/dobratz/01dobratz.html)
- Giles, C.L., Bollacker, K. and Lawrence, S. (1998), "CiteSeer: an automatic citation indexing system", in *Proceedings of the 3rd ACM Conference on Digital Libraries (DL98)*, Pittsburgh, PA, June 23-26, pp. 89-98, available at: <http://citeseer.ist.psu.edu/108208.html>
- Harnad, S. and Brody, T. (2004), "Comparing the impact of Open Access (OA) vs. Non-OA articles in the same journals", *D-Lib Magazine*, Vol 10 No. 6, available at: [www.dlib.org/dlib/june04/harnad/06harnad.html](http://www.dlib.org/dlib/june04/harnad/06harnad.html)
- Jacsó, P. (2006a), "Open access to scholarly indexing/abstracting information", *Online Information Review*, Vol. 30 No. 4, pp. 461-8.
- Jacsó, P. (2006b), "Open access scholarly databases – a bird's eye view of the landscape", available at: [www.ifla.org/IV/ifla72/papers/157-Jacso-en.pdf](http://www.ifla.org/IV/ifla72/papers/157-Jacso-en.pdf)
- Jacsó, P. (2006c), "Deflated, inflated, and phantom citation counts", *Online Information Review*, Vol. 30 No. 3, pp. 297-309.
- Kirsop, B. and Chan, L. (2005), "Transforming access to research literature for developing countries", *Serials Review*, Vol. 31 No. 4, pp. 246-55.
- Lawrence, S. (2001), "Free online availability substantially increases a paper's impact", *Nature*, available at: [www.nature.com/nature/debates/e-access/Articles/lawrence.html](http://www.nature.com/nature/debates/e-access/Articles/lawrence.html)
- OpCit (2006), "The effect of open access and downloads ('hits') on citation impact: a bibliography of studies", available at: <http://opcit.eprints.org/oacitation-biblio.html>
- Papin-Ramcharan, J. and Dawe, R.A. (2006), "Confronting the cost of information for a research library in the developing world – the University of the West Indies, Trinidad and Tobago's experience", *The International Information & Library Review*, Vol. 38 No. 1, pp. 15-24.

### Corresponding author

Peter Jacsó can be contacted at: [jacso@hawaii.edu](mailto:jacso@hawaii.edu)